1Л

УДК 004.89

В.В. Литвин, Р.Р. Даревич, Д.Г. Досин, А.С. Мельник, А.О. Мазур
Національний університет «Львівська політехніка», м. Львів, Україна
Фізико-механічний інститут ім. Г.В. Карпенка НАН України, м. Львів, Україна
vasyll@ukr.net, darevych@ipm.lviv.ua, dmytro.dosyn@gmail.com, ania@ipm.lviv.ua

Пошук релевантних прецедентів на основі адаптивних онтологій

У цій статті розглядається загальний підхід до функціонування інтелектуальних систем підтримки прийняття рішень на основі адаптивних онтологій, які використовують прецеденти. Розроблено математичний апарат для таких систем, який апробовано у задачі діагностування ревматологічних захворювань.

Вступ. Постановка проблеми у загальному вигляді

Технологія інтелектуальних систем підтримки прийняття рішень (ІСППР) є одним із найрозвинутіших напрямків штучного інтелекту. Дослідження у цій галузі полягають у розробці автоматизованих інформаційних систем, які застосовуються у тих сферах діяльності людини, які вимагають логічного міркування, певної майстерності та досвіду.

Сучасний рівень розвитку ІСППР відбувається у двох напрямках розроблення інтелектуальних агентів (ІА) [1]:

ІА, засновані на прецедентах (англійською – Case-Based Reasoning, або CBR);

ІА планування діяльності (пошук у просторі станів).

Практично всі ранні ІСППР моделювали хід ухвалення рішення експертом як дедуктивний процес з використанням логічного виведення, заснованого на правилах. Проте з часом було усвідомлено, що дедуктивна модель моделює один з найбільш рідкісних підходів, який використовує експерт під час вирішення проблеми. Існує багато слабоформалізованих завдань, для яких, можливо, будуть знайдені рішення. Крім того, існує ряд завдань, для яких ніколи не буде знайдено формального рішення (судочинство, медицина). Актуальність проблеми обумовлена і чисельністю таких завдань і практичною потребою знайти хоча б одне яке-небудь відповідне рішення там, де через відсутність строгого формалізованого методу не можна знайти всі або найоптимальніші рішення.

Насправді, замість того, щоб вирішувати кожну задачу, виходячи з первинних принципів, експерт часто аналізує ситуацію в цілому і згадує, які рішення ухвалювалися раніше в подібних ситуаціях. Потім він або безпосередньо використовує ці рішення, або, за необхідності, адаптує їх до обставин, що змінилися для конкретної проблеми.

Моделювання такого підходу до вирішення проблем, заснованого на досвіді минулих ситуацій, привело до появи технології логічного виведення, заснованого на прецедентах, і надалі – до створення програмних продуктів, що реалізовують цю технологію. Метод виведення за прецедентами ефективний, коли основним джерелом знань про задачу є досвід, а не теорія; рішення не є унікальними для конкретної ситуації, а можуть бути використані в інших випадках; метою розв’язування задачі є отримати не гарантований правильний розв’язок, а кращий з можливих. Виведення, засноване на прецедентах, є методом побудови ІСППР, які приймають рішення щодо даної проблеми або

ситуації за наслідками пошуку аналогій, що зберігаються в базі прецедентів [2]. Такий прецедент називають релевантним. З математичної точки зору серед елементів множини прецедентів $\mathrm{Pr} = \{\mathrm{Pr}_1, \mathrm{Pr}_2, \ldots, \mathrm{Pr}_N\}$ релевантним Pr_k є прецедент, для якого відстань до поточної ситуації S є найменшою, тобто

$$\mathrm{Pr}_k = \arg\min_i d(\mathrm{Pr}_i, S).$$

Задачу пошуку релевантних прецедентів можна вважати задачею класифікації, де класами є прецеденти. А поточну ситуацію необхідно віднести до певного класу.

Аналіз останніх досліджень та публікацій

Виведення на основі прецедентів – це метод прийняття рішень, в якому використовуються знання про попередні ситуації або випадки (прецеденти). Під час розгляду нової проблеми (поточного випадку) відшукується схожий прецедент як аналог. Прецедент – це опис проблеми або ситуації в сукупності з докладним вказанням дій, що виконуються в певній ситуації або для вирішення певної проблеми. Опис проблеми повинен містити всю інформацію, необхідну для досягнення мети виведення (вибору найбільш відповідного рішення). Наприклад, якщо мета полягає в діагностиці захворювань, то описова інформація повинна містити симптоми хворого, результати лабораторних досліджень.

Після того, як прецеденти знайдені, треба вибрати «найбільш відповідний» з них. Це визначається порівнянням ознак поточного випадку і вибраних прецедентів. Визначення методу, на якому ґрунтуватиметься знаходження міри схожості прецедентів, вирішується під час створення системи її розробниками. Найбільш популярним і часто використовуваним є метод «найближчого сусіда» (nearest neighbour). У його основі лежить той чи інший спосіб вимірювання ступеня близькості прецедента і поточного випадку за кожною ознакою (чи це текстова, числова або булева), яку користувач визнає корисною для досягнення мети.

Кажучи строгіше, вводиться метрика на просторі всіх ознак. У цьому просторі визначається точка, відповідна поточному випадку, і в рамках цієї метрики знаходиться найближча до неї точка серед точок, що відображають прецеденти. Зазвичай прогноз робиться на основі декількох найближчих точок, а не одної (K-nearest neighbours). Такий метод стійкіший, оскільки дозволяє згладити окремі випадки, випадковий шум, який завжди присутній в даних.

Кожній ознаці призначають вагу, що враховує її відносну цінність. Повністю ступінь близькості прецеденту за всіма ознаками можна обчислити, використовуючи узагальнену формулу вигляду:

$$\sum_k w_k \cdot sim\left(x_{ki}, x_{kj}\right), \quad \sum_k w_k = 1,$$

де w_k – вага k-ї ознаки, sim – функція подібності (метрика), x_{ki} і x_{kj} – значення ознаки x_k для поточного випадку i та j-го прецеденту, відповідно. Після обчислення ступенів близькості всі прецеденти ранґуються.

Вибір метрики (або міри близькості) вважається центральним моментом, від якого вирішальним чином залежить пошук релевантних прецедентів. У кожній конкретній задачі цей вибір відбувається по-своєму, з урахуванням головних цілей дослідження, фізичної і статистичної природи використовуваної інформації тощо. Як методи розв'язання таких задач використовуються алгоритми типу Lazy-Learning [3], зокрема відомі алгоритми найближчого сусіда і k-ближніх сусідів, нейронні мережі, генетичні алгоритми, байєсівські мережі [4], дерева рішень [5].

Основним недоліком нейромережної парадигми є необхідність мати дуже великий об'єм навчальної вибірки. Інший істотний недолік полягає в тому, що навіть натремована нейронна мережа є чорним ящиком. Знання, зафіксовані як ваги декелькох сотень міжнейронних зв'язків, абсолютно не піддаються аналізу й інтерпретації людиною, а відомі спроби дати інтерпретацію структурі налаштованої нейромережі виглядають непереконливими.

Популярність дерев рішень пов'язана з наочністю і зрозумілістю. Але для них дуже гостро стоїть проблема значущості. Річ у тому, що окремим вузлам на кожному новому побудованому рівні дерева відповідає все менша і менша кількість записів даних – дерево дробить дані на велику кількість окремих випадків. Чим їх більше, тим менше навчальних прикладів потрапляє в кожний такий окремий випадок, тим менш правильною буде їх класифікація. Якщо побудоване дерево дуже «кущисте» – складається з невиправдано великої кількості дрібних гілочок, – воно не даватиме статистично обґрунтованих відповідей. Як показує практика, в більшості систем, що використовують дерева рішень, ця проблема не знаходить задовільного рішення. Крім того, загальновідомо, і це легко показати, що дерева рішень дають корисні результати тільки у разі незалежних ознак. Інакше вони лише створюють ілюзію логічного виведення.

Генетичні алгоритми теж мають ряд недоліків. Критерій відбору хромосом і використовувані процедури є евристичними і далеко не ґарантують знаходження «кращого» рішення. Як і в реальному житті, еволюцію може «заклинити» на якій-небудь непродуктивній гілці. І, навпаки, можна навести приклади, як двоє неперспективних батьків, які будуть виключені з еволюції генетичним алгоритмом, виявляються здатними породити високоефективного нащадка. Це особливо стає помітним при вирішенні високорозмірних завдань зі складними внутрішніми зв'язками. Сама постановка завдання в термінах генетичних алгоритмів не дає можливості проаналізувати статистичну значущість отримуваного за їх допомогою рішення. Крім того, ефективно сформулювати завдання, визначити критерій відбору хромосом під силу тільки фахівцеві. Через ці чинники сьогодні генетичні алгоритми треба розглядати швидше як інструмент наукового дослідження, ніж як засіб аналізу даних для практичного застосування.

На нашу думку, позбутися ряд у вищенаведених недоліків дозволяють онтології предметної області та онтології задач.

Формування цілей. Розробити підхід для оцінювання релевантності прецедентів на основі онтологій. Побудувати метрику для здійснення такої оцінки. Апробувати цю метрику під час функціонування прикладної ІСППР.

Поняття адаптивної онтології

Формально онтологія складається з термінів (понять, концептів), організованих в таксономію, їх визначень і атрибутів, а також пов'язаних з ними аксіом і правил виведення. Тому під моделлю онтології O розуміють трійку вигляду:

$$O = \langle C, R, F \rangle,$$

де C – поняття, R – відношення між поняттями, F – інтерпретація понять та відношень (аксіоми). Аксіоми встановлюють семантичні обмеження для системи понять та відношень.

Для того, щоб можна було на онтологіях будувати метрику, нами запропоновано розширити цю модель за рахунок введення двох скалярних величин – ваги важливості понять та зв'язків [6], [7]. Коефіцієнт важливості поняття (зв'язку) – це чисельна міра, котра характеризує значущість певного поняття (зв'язку) у конкретній предметній області і динамічно змінюється за певними правилами у процесі експлуатації системи.

Отже,

$$O = \langle C, R, F, W, L \rangle,$$

де W – важливість понять C, L – важливість відношень R.

Визначену таким чином онтологію будемо називати адаптивною, тобто такою, що адаптується до предметної області (ПО) за рахунок модифікації понять та коефіцієнтів важливості цих понять і зв'язків між ними. Така онтологія однозначно представляється у вигляді зваженого концептуального графа (КГ) [8].

Побудова метрики для прецедентів у просторі ознак

Очевидно, що залежно від прецеденту ваги понять різні. Тобто насправді W – вектор вимірності кількості прецедентів $W = (W_1, W_2, \ldots, W_N)$. Надалі будемо розглядати лише один прецедент, тобто нижній індекс у вазі понять опускатимемо.

Побудуємо метрику для пошуку релевантних прецедентів на основі адаптивних онтологій. Нехай множина прецедентів $\mathrm{Pr} = \{\mathrm{Pr}_1, \mathrm{Pr}_2, \ldots, \mathrm{Pr}_N\}$ описується характеристиками (властивостями) $X = \{x_1, x_2, \ldots, x_M\}$. D_i – домен властивості x_i, w_{i_l} – коефіцієнт важливості властивості x_{i_l} прецедента Pr_i. Значення властивості x_i позначатимемо $z_i = z(x_i)$. Отже,

$$\mathrm{Pr}_i \leftrightarrow X_i = \left\{ x_{i_1} = z_{i_1}, x_{i_2} = z_{i_2}, \ldots, x_{i_k} = z_{i_k} \right\}, \text{ де } z_{i_j} \in D_{i_j}.$$

Позначимо I_i – множину індексів властивостей прецедента Pr_i. Тоді відстань між прецедентом Pr_i та поточною ситуацією S визначається як:

$$d_i = \sum_{i_l \in \overline{I}_i} \varphi\left(z_{i_l}, z_{i_l}^S \right), \tag{1}$$

де z_{i_l} – значення властивості x_{i_l} прецедента Pr_i, $z_{i_j}^S$ – значення властивості x_{i_l} поточної ситуації S, $\overline{I}_i$ – множина індексів важливих властивостей прецедента Pr_i, $\overline{I}_i = \overline{I}_{i1} \cup \overline{I}_{i2} \cup \ldots \cup \overline{I}_{iN_i}$, N_i – кількість властивостей, які необхідно розглянути, щоб прийняти рішення стосовно прецедента Pr_i. Тобто

$$\overline{I}_{i1} = \left\{ i_{s1} \,\middle|\, i_{s1} = \arg\max_{i_l \in I_i} w_{i_l} \right\}, \quad \overline{I}_{i2} = \left\{ i_{s2} \,\middle|\, i_{s2} = \arg\max_{i_l \in I_i / i_{s1}} w_{i_l} \right\}, \quad \overline{I}_{i3} = \left\{ i_{s3} \,\middle|\, i_{s3} = \arg\max_{i_l \in I_i / i_{s1} / i_{s2}} w_{i_l} \right\}, \ldots$$

Розглянемо функцію $\varphi(\xi, \eta)$. Очевидно, що ξ може бути діапазоном, тобто нечіткою підмножиною $\xi \subseteq D$, де D – універсальна множина; числовим значенням або нечисловим значенням. Залежно від цього $\varphi(\xi, \eta)$ визначається по своєму, а саме:

$$\varphi(\xi, \eta) = \begin{cases} 1 - \mu_\xi(\eta), \ \xi - \text{нечітка множина}, \\ \lambda \cdot |\xi - \eta|, \ \xi, \eta - \text{числові значення}, \\ 1 - \mu(\xi, \eta), \ \xi, \eta - \text{нечислові значення}, \end{cases} \tag{2}$$

де $\mu_\xi(\eta)$ – коефіцієнт впевненості того, що η належить нечіткій підмножині ξ; λ – числова величина, яка залежить від ПО, щоб добуток $\lambda \cdot |\xi - \eta| \in [0,1]$ (розмірність величини λ обернено-пропорційна до розмірності величин ξ та η, тобто, якщо ξ та η є маса і вимірюється в кг, то λ вимірюється в кг$^{-1}$); $\mu(\xi, \eta) \in [0,1]$ – нечітка величина подібності значень ξ та η. Наприклад $\mu(\xi, \eta) = 1$, якщо $\xi = \eta$, $\mu(\xi, \eta) = 0{,}9$, якщо $\xi \approx \eta$, $\mu(\xi, \eta) = 0$, якщо $\xi \neq \eta$.

Тепер постає задача – як визначити ваги понять та відношень. Для цього нами пропонується використовувати методи інтелектуального аналізу даних (ІАД), наведені вище в аналізі останніх досліджень, наприклад дерева рішень. Нами пропонується використовувати методи ІАД для визначення ваг деякої підмножини понять. Такі ваги називатимемо базовими. Потім на основі онтології ПО розвинути отримані ваги на всю онтологію. Цю процедуру здійснити для кожного прецедента. Тоді для пошуку релевантного прецедента використовувати значення тих N_i понять, які для відповідного прецедента мають найбільші ваги. Що стосується ваг важливості відношень, то нами пропонується задати їх такими, як це наведено у табл. 1. Групи відношень взяті із [9]. Вважаємо, що вага вертикальних зв'язків (ієрархія, агрегація) рівна 1,2 (чим конкретніше, тим краще). Квантатативні відношення не розглядаються, оскільки синонімія та королювання ніяк не впливають на значення ознак. Відразу вважається, що це одна і та ж ознака.

Таблиця 1 – Ваги важливості відношень

Група відношень	Відношення	Значення ваг важливості L
Ієрархія	Рід↔вид	*1,2*
	Ознака↔значення ознаки	*1,2*
	Інваріант↔варіант	*1,2*
Агрегація	Ціле↔частина	*1,2*
	Об'єкт↔простір реалізації (локалізації) об'єкта	*1,2*
	Об'єкт↔властивість/ознака	*1,2*
	рівень↔одиниця рівня	*1,2*
Функціональні	Об'єкт дії↔дія↔суб'єкт ді ї	*1*
	причина↔наслідок	*0,9*
	умова↔дія	*0,9*
	явище↔дія	*0,9*
	стан↔дія	*0,9*
	явище↔стан	*0,9*
	інструмент↔дія	*0,9*
	дані↔дія	*0,9*
	дані↔величини	*0,9*
Семіотичні	Термін↔спосіб вираження	*0,2*
	Термін↔спосіб подання	*0,2*
	Термін↔метазнак терміну	*0,2*

Отже, нами пропонується загальний підхід для знаходження релевантних прецедентів, який складається із 3 кроків:

Задати базові ваги важливості понять на основі ІАД.

Розповсюдити ці ваги на всю онтологію.

Використовуючи (1) – (2), визначити релевантні прецеденти.

Розглянемо перші два кроки детальніше. Нехай для розв'язування задачі класифікації нами побудовано дерево рішень (ДР). Вершини (ознаки) цієї гілки знаходяться на k рівнях. Очевидно, що чим вищий рівень, тим більш значуща ознака, яка на цьому рівні знаходиться. Ця евристична думка має бути відображена в значеннях ваг

цих ознак. Крім того, пропонується ці ваги нормувати, тобто щоб їх сума для кожного прецедента (гілки) була рівна 1.

Розглянемо два способи визначення ваг базових ознак, які задовольняють вище–описані два припущення.

1. Арифметичний спосіб. Визначаються як відношення різниці (k+1) рівня дерева та рівня, на якому знаходиться ознака до суми всіх рівнів гілки, тобто базуються на сумі арифметичної прогресії:

$$w_i = \frac{k+1-i}{\sum_{j=1}^{k} j} = \frac{k+1-i}{\frac{(1+k)k}{2}}.$$

2. Геометричний спосіб. Базуються на сумі геометричної прогресії. Ваги понять, які лежать на i-му рівні ДР, визначаються таким чином:

$$w_i = \frac{2^{k-i}}{2^k - 1}.$$

Отриману множину базових ваг позначимо W_B. Тепер необхідно їх розвинути на всю онтологію ПО, використовуючи таксономію понять онтології, відношення між поняттями та їх інтерпретацію. Математично (формально) цей процес запишемо у вигляді:

$$W_B \xrightarrow{O} W. \quad (3)$$

Для ознак, які беруть участь у ДР, до ваг важливості понять відповідного прецедента додаємо вагу, отриману на основі ДР. Потім розглядаємо функціональні та семіотичні зв’язки, якщо вони визначені в онтології. Квалітативні функціональні відношення діляться на симетричні R_S (деяка підмножина горизонтальних зв’язків) та несиметричні R_N (вертикальні зв’язки, інша підмножина горизонтальних зв’язків). Те, що елементи містять між собою симетричний зв’язок, позначатимемо подвійною стрілкою $C_i \leftrightarrow C_j$, несиметричний – одинарною від області визначення в множину значень $C_i \rightarrow C_j$.

Очевидно, що елементи, які беруть участь у симетричних зв’язках, є рівносильними. Тому ваги L симетричних відношень рівні одиниці. Таким симетричним зв’язком є «Об’єкт дії↔дія↔суб’єкт дії», який належить до групи функціональних зв’язків. Всі інші функціональні зв’язки є близькими до симетричних, тому їх ваги рівні 0,9. Отже, якщо відома вага W_i терміну C_i і цей термін має симетричний зв’язок із терміном C_j, вага якого невідома, то $W_j = L \cdot W_i$. Для несиметричних зв’язків $C_i \rightarrow C_j$ отримаємо аналогічне співвідношення

$$W_j = L \cdot W_i, \quad (4)$$

якщо відомо вагу C_i, і

$$W_i = \frac{W_j}{L}, \quad (5)$$

якщо відомо вагу C_j. Семіотичні відношення мають вагу 0,2.

ІСППР діагностування ревматологічних захворювань

Розглянемо приклад функціонування ІСППР у галузі медицини, а саме ревматології. Як прецеденти нами було розглянуто сім захворювань: Pr = {‘Ревматоїдний

артрит (РА)’, ‘Деформуючий остеоартроз (ДО)’, ‘Системний червоний вовчак (СЧВ)’, ‘Анкілозуючий спондиліт (АС)’, ‘Гостра ревматична лихоманка (ГРЛ)’, ‘Подагричний артрит (ПА)’, ‘Системна склеродермія (СС)’}. Нами виділено 27 ознак, які впливають на ці захворювання. Наведемо деякі з них: $X=\{$*Симетричний поліартрит дрібних суглобів кистей рук, Звуження суглобової щілини, Остеофіти, Біль у суглобах при фізичному навантаженні, Феномен Рейно, Експресія HLA-D27 антигену, Виявлення LE-клітини у крові*$\}$. Очевидно, що не всі ознаки присутні в окремо взятому захворюванні. Тому залежно від прецеденту важливість цих ознак різна. Саме таку вагу важливості нам необхідно визначити, щоб для діагнозу використати формулу (1). Детальніше домени ознак розглянемо, коли визначатимемо вагу їх важливості.

Використовуючи різні ревматологічні довідники, ми побудували онтологію ревматології. В онтології відображено взаємозв’язок між ознаками, їх вплив на захворювання залежно від значення ознаки. Використовуючи цю онтологію, визначимо вагу важливості цих ознак залежно від захворювання. Насамперед обчислимо початкові вагові коефіцієнти. Для цього побудуємо дерево рішень (ДР) на основі архіву даних про захворювання колишніх пацієнтів. Отримане ДР наведено на рис. 1. Щоб спростити виклад, присвоїмо кожній ознаці індекс, який відповідає її коду в базі даних. На рис. 2 наведено структуру БД, деякі дані беруться із онтології та дерева рішень. Для задання ваг базових ознак прецедентів скористаємось арифметичним способом їх визначення. Для РА отримаємо такі значення: $W_{12}^0=\frac{1}{2}$, $W_1^0=\frac{1}{3}$, $W_7^0=\frac{1}{6}$. Верхній індекс (0) вказує, що ваги є початковими. До уваги беруться лише ознаки, які є наявними у цьому захворюванні. Тому, наприклад для АС, глибина ДР рівна двом і ми отримуємо початкові ваги лише для ознак «*Біль у суглобах в спокої*» (код в БД 14) та «*Сакроілеїт*» (код – 5). Їх ваги для АС будуть $W_{14}^0=\frac{2}{3}$ та $W_5^0=\frac{1}{3}$. Для ПА та СС взагалі лише одна базова ознака. Так для ПА – це «*Тонуси*» (код – 27), а для СС – «*Феномен Рейно*» (код – 17). Початкові ваги цих ознак для відповідних прецедентів рівні одиниці.

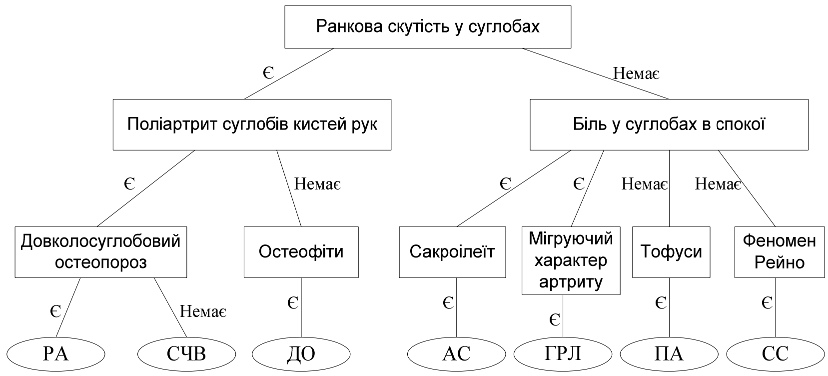


Рисунок 1 – Дерево рішень, отримане на основі аналізу даних ревматологічних захворювань

Розглянемо процес розвинення ваг базових ознак на всю онтологію ревматології, використовуючи (4) та (5). Для цього проаналізуємо нашу онтологію. Онтологія реалізована в редакторі Protégé-OWL_4.1. На верхньому рівні ієрархії онтології знаходяться ознаки та прецеденти (рис. 3 а, б). Ієрархію відношень представимо у вигляді, розгля–нутому в 3-му розділі. Отже, в онтології ревматології існує 6 типів відношень (рис. 3 а). Перші два типи (ієрархія, агрегація), як правило, задаються у вигляді ієрархії понять, однак можливі випадки, що існуватиме множинне наслідування, або деяке поняття буде

складовим кількох інших понять. Тоді ці відношення необхідно визначити в закладці Object Properties.

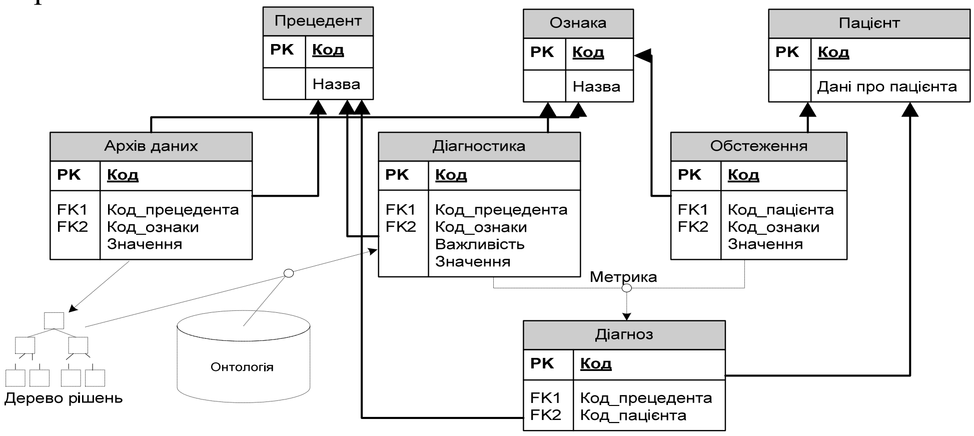
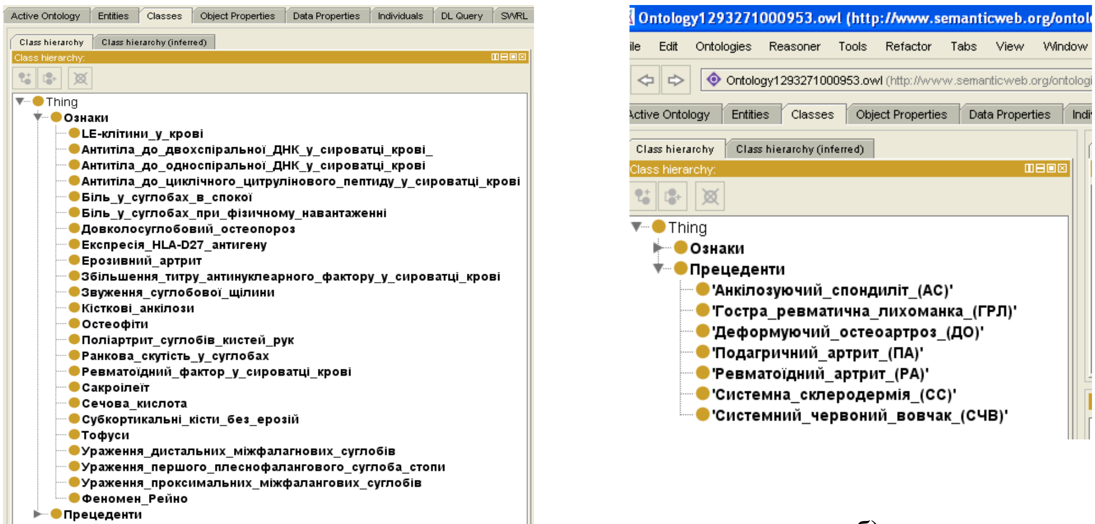


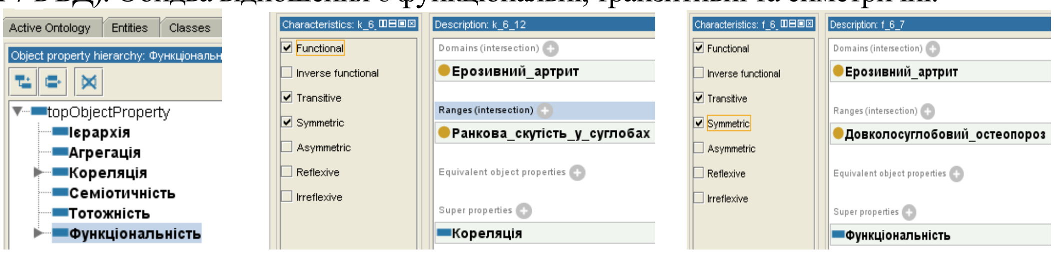
Рисунок 2 – Структура бази даних ІСППР діагностування ревматологічних захворювань



а) б)

Рисунок 3 – Ієрархія понять онтології ревматології

Однак основними для перерахунку ваг важливості понять онтології є функціональні та кореляційні відношення. Окремі з них наведено на рис. 4 а) – в). В обидвох відношеннях доменом є ознака «Ерозивний артрит» (ключ 6 в БД). Множиною значень у першому випадку є «Ранкова скутість у суглобах» (ключ 12 в БД) та «Довколосуглобовий остеопороз» (ключ 7 в БД). Обидва відношення є функціональні, транзитивні та симетричні.



а) б) в)

Рисунок 4 – Види відношень онтології ревматології

Перерахувавши ваги важливості понять згідно з формулами (4) – (5) отримаємо такі ознаки для різних захворювань, які необхідно використати під час обчислення відстаней за формулою (1) (табл. 2). Ми взяли не більше як три ознаки для кожного прецедента. У випадку, якщо вага 2-го або 3-го поняття була менша за 0,5, то відповідно до розгляду бралась одна або дві ознаки (так як у випадку ПА та СС). Окрім того, використання такої властивості, як Disjoint (тобто заперечення) в онтології ревматології привело до того, що відсутність ознаки також має значну вагу (такий результат ми отримали у випадку ГРЛ, коли відсутній «Ерозивний артрит»).

Таблиця 2 – Залежність захворювань від ознак

Прецедент	Ознака 1	Ознака 2	Ознака 3
Ревматоїдний артрит (РА)	Поліартрит суглобів кистей рук	Ерозивний артрит	Довколосуглобовий остеопороз
Деформуючий остеоартроз (ДО)	Остеофіти	Субкортикальні кісти без ерозій	Біль у суглобах при фізичному навантаженні
Системний червоний вовчак (СЧВ)	Біль у суглобах у спокої	Антитіла до двоспіральної ДНК у сироватці крові	Поліартрит суглобів кистей рук
Анкілозуючий спондиліт (АС)	Сакроілеїт	Біль у суглобах у спокої	Експресія HLA-D27 антигену
Гостра ревматична лихоманка (ГРЛ)	Моно-олігоартрит великих суглобів	Мігруючий характер артриту	Ерозивний артрит (*відсутність*)
Подагричний артрит (ПА)	Тофуси	Сечова кислота	
Системна склеродермія (СС)	Феномен Рейно		

Аналіз табл. 2 порівняно з ДР, яке наведене на рис. 1, показує, що для ПА та СС ознаки практично не змінились. Для інших захворювань помінялись в одній або двох позиціях. Перевірка реальних захворювань показала, що використання ДР правильно класифікує захворювання в 64% випадках, а класифікація отримана згідно з онтологією ревматології в 79% випадках.

Висновки

Розроблено математичну модель функціонування інтелектуальних систем підтримки прийняття рішень на основі адаптивних онтологій для пошуку релевантних прецедентів. Ця модель базується на метриці. Для побудови такої метрики використано адаптивну онтологію. З цією метою у загальноприйнятий триелементний кортеж, який задає онтологію (множина понять, відношень та їх інтерпретація), нами додано дві скалярні величини (важливість понять та відношень), які використовуються для обчислення необхідних відстаней. Розглянуто способи задання початкових коефіцієнтів важливості понять та зв'язків на основі інтелектуального аналізу даних та їх розповсюдження на всю онтологію. Розроблено загальний підхід функціонування ІСППР, які використовують розроблену модель, який складається із трьох кроків. Розглянуто приклад функціонування такої системи для діагностики ревматологічних захворювань. Отримані результати показують ефективність розробленої моделі порівняно з деревами рішень.

Література

1. Каменнова М.С. Корпоративные информационные системы: технологии и решения / М.С. Каменнова // Системы Управления Базами Данных. – 1995. – № 3. – С. 88-99.
2. Funk P. Advances in Case-Based Reasoning / P. Funk, P.A. González-Calero // 7th European Conference, ECCBR 2004. – Madrid, Spain. – P. 375-380.
3. Wettschereck D. A Review and Empirical Evaluation of Feature Weighting Methods for a Class of Lazy Learning Algorithms / D. Wettschereck D. Aha, T. Mohri // Artificial Intelligence Review. – № 11. – P. 273-314.
4. Рассел С. Искусственный интеллект / С. Рассел, П. Норвиг. – М. ; С.-П. ; К. : Вильямс, 2006. – 1408 с.
5. Цветков А.М. Разработка алгоритмов индуктивного вывода с использованием деревьев решений / А.М. Цветков // Кибернетика и системный анализ. – № 1. – 1993. – С. 174-178.
6. Інтелектуальні системи, базовані на онтологіях / Д.Г. Досин, В.В. Литвин, Ю.В. Нікольський, В.В. Пасічник. – Львів : Цивілізація, 2009. – 414 с.
7. Даревич Р. Р. Підвищення ефективності інтелектуального аналізу тексту шляхом зважування понять в моделі онтології / Д.Г. Досин, В.В. Литвин, З.Т. Назарчук // Искуственный интеллект. –2006. – № 3. – С. 500-509.
8. Sowa J. Conceptual graphs for a database interface / J. Sowa // IBM Journal of Research and Development. – 1976. – Vol. 20, № 4. – P. 336-357.
9. Найханова Л.В. Технология создания методов автоматического построения онтологий с применением генетического и автоматного программирования / Л.В. Найханова. – Улан-Удэ : Издательство БНЦ СО РАН, 2008. – 244 с.

Literatura

1. Kamennova M.S. Database Management Systems. № 3. 1995. P. 88-99.
2. Funk P. 7th European Conference, ECCBR. Madrid. 2004. P. 375-380.
3. Wettschereck D. Artificial Intelligence Review. № 11.P. 273-314.
4. Russell S. Moscow-S.-Ptb.: Williams. 2006. 1408 p.
5. Tsvetkov A.M. Cybernetics and System Analysis. № 1. 1993. P. 174-178.
6. Dosin D.G. Lviv: Tsivilizatsiya. 2009. 414 p.
7. Dosin D.G. Artificial Intelligence. IPAI "Nauka i Osvita" 2006.Issue. 3. P. 500-509.
8. Sowa J. IBM Journal of Research and Development. Vol. 20. № 4. 1976. P. 336-357.
9. Naykhanova L.V. Ulan-Ude: BSC SB RAS Publisher. 2008. 244c.

В.В. Литвын, Р.Р. Даревыч, Д.Г. Досын, А.С. Мельнык, А.О. Мазур

Поиск релевантных прецедентов на основе адаптивных онтологий

В статье рассматривается общий подход к функционированию интеллектуальных систем поддержки принятия решений, базирующихся на адаптивных онтологиях, использующие прецеденты. Разработан математический аппарат для таких систем, апробирован в задаче диагностирования ревматологических заболеваний.

V.V. Lytvyn, R.R. Darevych, D.G. Dosyn, A.S. Melnyk, A.A. Mazur

Searching for Relevant Precedent Based on Adaptive Ontology

In this article the general approach to the functioning of intelligent decision support systems based on adaptive ontologies using precedents is described. A mathematical formalism for such systems was tested in the task of diagnosing rheumatologic diseases.

Стаття надійшла до редакції 07.04.2011.